DOCUMENT RESUME

ED 368 175 FL 021 860

AUTHOR Brown, Annie
TITLE LSP Testing: The Role of Linguistic and Real-World Criteria.
PUB DATE Apr 93
NOTE 11p.; Paper presented at the Annual Meeting of the Southeast Asian Ministers of Education Organization Regional Language Center Seminar (28th, Singapore, April 19-21, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluation Criteria; Foreign Countries; Higher Education; Japanese; *Language Proficiency; *Languages for Special Purposes; *Language Tests; *Licensing Examinations (Professions); Oral Language; *Test Construction; Test Format; Test Items; Test Validity; Tourism; Verbal Tests
IDENTIFIERS Australia

ABSTRACT

Issues in testing proficiency in languages for special purposes (LSP) are examined in the context of the development of an advanced oral test in Japanese for tour guides. The test, designed at an Australian university for use in the Australian tourism industry, was to be designed to evaluate both language proficiency and skills in appropriate guide-client interaction. The major issue involves the use of both language specialists and profession-specific specialists in development of test items and assessment criteria. It was found, during construction of the test, that attention to definition of real-world, non-linguistic criteria for assessment is essential, and that the best way to ensure test relevance is to involve representatives of the industry at all stages of test development. (Contains 15 references.) (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

LSP testing: the role of linguistic and real-world criteria*

Annie Brown

The University of Melbourne

ED 368 175

1. Introduction

Specific purpose language tests commonly attempt to reflect the future context of language use by simulating the criterion performance in the test tasks. Thus their major feature could be said to be their predictive value. However, as assessments of such performances tend to be made by language specialists, rather than representatives of the profession itself, they inevitably focus on purely linguistic criteria. It could be argued that such assessments may not adequately predict the test-taker's ability to perform in the occupational context, where linguistic skills are but one factor in successful performance, and it may then be appropriate to include both assessments of purely linguistic competence *and* assessments involving a perception of professional competence (using both linguistic and 'real-world' criteria). This would necessarily involve representatives of the profession in question being involved at all stages of the test development - the needs analysis, item writing and development of the assessment criteria. It will be argued that in the development of specific purpose language tests, such experts play a crucial role in all aspects of the test development process, and can add substantially to our perceptions as language testers of those features of successful occupational communication which are vital in ensuring construct and content validity.

In this paper these issues will be explored in relation to the development of an advanced level occupation-specific oral language test, the Japanese Language Test for Tour Guides. The context in which the test was to be introduced (that is, as part of an industry-driven accreditation scheme for tour guides) required that the test be designed to evaluate not only language proficiency levels, but whether the candidates are able to interact 'appropriately' in guide-client interaction.

The NLLIA Language Testing Centre at the University of Melbourne was commissioned to develop oral testing procedures to measure the Japanese language skills of Australian Japanese-speaking tour guides as part of the industry-driven development of accreditation procedures. This work was funded by DEET, in part through the Japanese Proficiency Project of the National Languages and Literacy Institute of Australia and in part through Tourism Training Australia. It was determined that the assessment procedures were to have a dual function: firstly to indicate to employers the language proficiency of entrants to the profession through optional certification, and secondly as a selection procedure for all applicants for the newly-developed TAFE Japanese tour guide training courses.

The test development process has two main aspects which are of relevance in this paper: firstly, the development of the specifications and the test item writing; and secondly, the development of the assessment criteria. To some extent, these two developments occurred in tandem, as the drafting of the specifications naturally included reference to the assessment procedures, as the specifications describe the constructs which the test purports to measure. However, at this stage the assessment criteria are embryonic and can only then be finalised once the test has been trialled and the test development team can confirm that these constructs are indeed measurable.

FL021860

2. Task development

* Paper presented at the SEAMEO-RELC Regional Seminar, Singapore, April 19-21, 1993

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Annie Brown
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

The Japanese Test for Tour Guides is a performance test, that is, a test in which the candidates are required to perform in a simulation of the actual target task. Such tests are well suited to such situations where the target situation is able to be clearly delineated and described, and where this is the case such a test should be employed. As Jones 1979 points out, 'It is impossible for a language test to predict task-oriented proficiency unless it includes or approximates actual samples of the tasks'.

In the development of the test tasks, the issues to be addressed include the role of industry representatives in identifying those types of interaction which are central to the work of a Japanese-speaking tour guide (the needs analysis) and in providing feedback on the authenticity of the interaction elicited within the tasks.

The first stage of the test development process involved an analysis of the types of speech acts and interactional patterns important in tour guide / tourist communication in the Japanese market. These were identified in two ways; firstly, permission was sought to accompany tours (conducted by both Japanese native speaker and non-native speaker guide) and either take notes or videotape the interaction for later analysis. Secondly, this live data was supplemented by reviewing the literature on Japanese in the tourism industry as well as conducting structured interviews with industry representatives (both guides and employers). On the basis of the information gathered, a list of the most salient types of interaction was drawn up. These were then fed back to a group of experts, consisting of representatives of the tour guiding profession (employers and guide trainers) and linguists with expertise in Japanese for tourism, for comment on the appropriateness and representativeness of the proposed tasks.

A decision had to be made regarding the number and type of interactions to include in the test, the range of speech acts. Shohamy (1992) refers to research which demonstrates that the type of interaction elicited in a language test can affect the test takers' scores. In other words, the level of performance is dependent on the type of language task. As she points out, it is therefore important to provide a range of tasks eliciting a variety of language discourse types in order to obtain a valid measure of the test taker's ability. Therefore, in the overall design of the test we decided to ensure that a range of different aspects of the work interaction were included, that the sampling was not simply representative but also broad.

This first stage, identifying the types of interaction which are required of guides and the contexts in which they typically occur (and which it is deemed appropriate to include in the test[1]), led to the following test structure (see Appendix 1 for details):

Phase 1: Introduction
Phase 2: Optional tours
Phase 3: Handling difficult situations
Phase 4: Cultural presentation
Phase 5: Giving instructions
Phase 6: Itinerary and tourist attraction

---

[1] One notable omission from the test was a measure of their ability to interpret, despite this being a frequently required skill. While interpreting is required of certain guides, especially those involved in specialist tours, it is opinion of many people within the profession that it should rightly be undertaken by professional interpreters. (There is a separate professional body with its own accreditation system, NAATI.) The fact that interpreting is required of guides is viewed with dissatisfaction by NAATI as well as by many people within the tour guiding profession, although for different reasons. NAATI is keen to ensure that the professional status of interpreters is maintained through ensuring that only NAATI qualified interpreters are employed, and the guiding representatives are of the opinion that guides who are required to interpret as part of their duties should be recognised and paid accordingly. Thus it was strongly argued that to include interpreting in the tour guide test, although a reflection of the actual situation, would be to validate it as a tour guiding role.

Once this format had been agreed upon, the next stage was to develop draft items for each phase. Five or six alternatives were to be developed for each. In the writing of the individual tasks we needed to ensure that each represented the criterion situation as accurately as possible in terms of the demands it placed on the test candidates. At this stage it had been determined that the reporting procedure was to be descriptive and should include reference not only to linguistic ability but also to the candidates' overall ability to fulfil the requirements of the task successfully. This decision was based on a consensus of opinion amongst the industry representatives involved in the test development, who considered that language should not (or could not) be divorced from other aspects of performance, that there was a 'swings and roundabouts' effect in guiding communication, whereby a guide with limited Japanese language skills might well be able to compensate for this through other traits. It was felt that where a test was being designed to involve test takers in simulations of the target performance, this would provide an opportunity to assess these other traits too, that is, overall performance on the task. These 'other traits' do not refer to occupational knowledge (as indeed it was important to ensure that this was not a feature, given that the test was to be taken both by guides with experience and by people hoping to enter the industry), rather they refer to general non-language based communication skills and traits which also affect the listener's evaluation of the quality of the performance. McNamara (1990), points out the role of such traits by distinguishing two types of language performance test, strong and weak.

> "The strong sense of the term is as follows: a second language performance test is a performance test in which language ability will only be one of many criteria used in assessing performance. Performance will primarily be judged on real world criteria, that is, the fulfilment of the task set. Such a test thus involves a second language as the medium of the performance; the performance itself (or rather, its outcomes) is the target of assessment. .....Adequate second language proficiency is a necessary but not sufficient condition of success on the performance task.....Performance of the communicative task (persuasion, reassurance, etc.) will be assessed against real-world criteria (am I persuaded? do I feel reassured?) and non-linguistic contextual factors such as the personality and sympathetic qualities of the person doing the persuading or reassuring will be involved in the assessment."

A weak performance test on the other hand is:

> "A test of second language performance: that is, performance on a task, the purpose of which is to elicit a language sample so that second language proficiency may be assessed."

So, in a weak performance test the content of the test tasks is to some extent merely providing face validity. There is no necessity for the sample to be realistic as linguistic features alone are the focus of the assessment. However, in this project the assessment was to address both linguistic and real-world aspects of the performance, that is, the test was to be a strong performance test. It was not enough that the test merely elicit a sample of language. Rather it was crucial to the construct validity of the test that it elicit a performance which reflected real life performance as closely as possible not just in terms of content, but also in terms of the cognitive strategies brought to bear on the interaction - in other words the interaction must be approached as far as possible in the same way as in real life - which means among other things making sure the candidates have sufficient background knowledge and are aware that the purpose of the interaction is more than just a demonstration of their linguistic skills.

It was clear, therefore, that the role of the industry informants would not stop at the test specification stage, as is often the case in LSP test development, but that their input would be required during the task development stage to ensure that the conditions surrounding the task were adequate to enable the candidates to demonstrate not only their linguistic skills but also their ability to fulfil the task in occupational terms. In other words, they needed to confirm that the tasks were capable of producing realistic interaction. Thus the nature and comprehensibility of the input, the

content of the test rubric and the contents of the test handbook (where the rationale of the test is laid out) required careful planning in order to ensure that the tasks were in fact meaningful to and feasible for both guides and non-guides in the same way, that they were able to approach the task as they would in the real-life context, and that they were not required to bring to the interaction cognitive processes which are not part of the real life interaction (guesswork, translation skills, speededness, focus on form at the expense of content, and so on). Issues to be considered included the language of the input, the provision of technical terms, the amout of preparation time, the amount of content provided, the instructions to the interviewer, and so on.

While it may seem trite to state this, it cannot be denied that much test development fails to address the issue of whether the test interaction reflects real life interaction. There often appears to be an assumption that the directness of an oral test, that is, the fact that it is a live face-to-face interaction, and the fact that the task design is based upon a detailed needs analysis and resultingly complex test specifications, will inevitably result in realistic tasks. In fact the testing situation itself imposes its own restrictions on the interaction, and these can limit the extent to which the test performance reflects real life performance. Authenticity of interaction should not be assumed, rather the onus is upon the test developer to ascertain this by gathering feedback both from experts in the occupational field and from the test-takers, the trial candidates. Only then can one say with confidence that the test reflects real-life interaction as closely as possible within a testing context .

In the development of the trial test items, feedback was gathered as follows:

1. Occupational representatives were asked whether they considered the items (including the rubric and the input, as well as the information and instructions which would be given to test interviewers) to be effective in enabling candidates (both with and without guiding experience) to produce naturalistic interaction. Modifications were made to the items and associated information/instructions as necessary.

2. Limited pre-trialling was undertaken and feedback was sought from the participants (again both with and without guiding experience) as to whether the interaction reflected how they would perform given the same situation in real-life. Modifications were made as necessary.

3. Further feedback was also gathered at a later stage from candidates taking part in the test trials. At this stage, feedback was also sought on the validity of the test as a whole. The following questions received responses (on a scale of 1 to 5, 5 being the most positive), which indicated that the test had a high degree of perceived validity, and that candidates without guiding experience did not feel disadvantaged:

To what extent do you think this test is appropriate for the assessment of oral language skills for tour guiding?
Guides (N=32) 4.6, Non-guides (N=16) 4.2

To what extent do you feel you were able to demonstrate the extent of your speaking ability adequately?
Guides 3.7, Non-guides 3.6

How did you react emotionally to the test as a whole? Guides 3.4, Non-guides 3.4

3. Assessment

In the tour guiding profession, language skills and occupational knowledge are only two of the features considered to be important in professional competence. Personality, maturity, presentation skills and appropriate intercultural behaviour are also considered to be crucial. As mentioned earlier, it was required of the test developers that the test assess not only language skills but also these other real-world aspects of professional competence. The industry experts, unlike many

language specialists who are often solely responsible for test design, did not see the two as clearly separable. Rather they were of the opinion that relevant, useful feedback on test taker performance must also include reference to real-world criteria, in other words, will the client be satisfied with the interaction (going back to the earlier definition of a strong test - 'am I persuaded? do I feel reassured?'). When, in the needs analysis stage, the industry representatives were asked to describe and comment on the range of tasks to be included in the test, they continually made reference to the quality of performance in these terms, where quality could not be quantified through assessment of language proficiency alone, and where different types of performance would require the demonstration of different abilities - the ability to demonstrate sympathy, to resolve difficult situations, to give advice, to promote activities, to give clear instructions and to describe events and places in a way which makes them attractive to the listener.

Successful communication, then, is not just a measure of the linguistic product, rather it involves the degree to which meaning and attitude (both verbal and non-verbal) intended by the speaker gets across to the listener. Olshtain and Blum-Kulka (1985) point out that an utterance is the performance of a communicative act which has an interactional function, and we need therefore to look not just at the form of the language but at the effect on the listener. It follows from this that in measuring authentic language use we need to look not just at the language but at whether the user is responding to the requirements of the listener. Therefore, we are not only concerned with the test candidates language skills, but also with their strategies for negotiating, obtaining and presenting meaning. The test candidates' ability to predict and respond to the listeners' needs goes well beyond mere second language ability.

Recent models of communicative competence, (Canale and Swain 1980, Canale 1983, Bachman 1990) recognise the fact that communication consists of more than simply knowledge of the language. These models of communicative competence owe a lot to the work of Hymes, who, in asserting that '"real" language performance involves linguistic as well as extra linguistic, social and psychological variables, all of which operate in constrant interaction' (Hymes 1971, quoted in Shohamy and Reves), defined the complexity of communication.

Canale (1983) refers to the place of non-linguistic variables in communication in his notion of strategic competence thus:

> "Strategic competence: mastery of verbal and non-verbal strategies both (a) to compensate for breakdowns in communication due to insufficient competence or performance limitations and (b) to enhance the rhetorical effect of utterances."

Strategic competence was, then, presumably what the industry representatives were referring to when they said that weak language skills could be compensated for by other means, that successful interaction in the tour guide context consisted of more than simply second language ability.

Bachman's model of communicative competence (1990, 1991) presented in his seminal text 'Fundamental Considerations in Language Testing', and perhaps currently the most comprehensive definition in terms of language testing, includes language knowledge, consisting of organisational knowledge (grammatical and textual knowledge) and pragmatic knowledge (illocutionary and sociolinguistic knowledge), knowledge of the world and strategic competence (the ability to utilise ones resources effectively in a communicative context). However, while he recognises that non-verbal as well as verbal strategies are "clearly an important part of strategic competence in communication", he chooses not to address them within his model.

Such models of communicative competence were found to be too limited in their scope to provide us with a framework for the development of the assessment procedures for the Tour Guide Test. Although they refer to the existence of non-linguistic variables, they do not address them in any depth. This is perhaps to be expected as they are models for the assessment of second *language*

ability and restrict themselves to language as it is understood in the weak test sense. However, it means they have little to offer for the assessment of communication in the strong sense.

It had been determined that there would be two aspects to the assessment of performance on the Tour Guide test:

1) second language proficiency, measured in terms of those linguistic criteria considered relevant to the context for which the assessments were being made;
2) overall communicative success in terms of how well the test taker fulfilled the task, according to as yet undefined criteria. It was assumed that this would not correlate fully with pure linguistic proficiency, as we know that in their first language some people are better at explaining instructions, are more interesting, are better able to persuade or to express sympathy than others, and so on, and that this is not necessarily a result of better language proficiency. We are assuming, then, that in a second language context, real-world success will reflect a combination of second language skills and these other interactional skills and that these will interact with each other in the evaluation of performance.

An expert group consisting of experienced Japanese language teachers, Japanese tourism skills trainers and tour co-ordinators and the three item developers was convened. These people met several times over two weeks to analyse the trial videos (of which we had 52) and to establish the assessment criteria and describe the levels of performance on each of these.

(a) The linguistic criteria
Through discussion with our informants we had determined areas of linguistic competence which were considered relevant within this context. These included fluency, resources of grammar and expression, appropriate level of politeness (including use of honorifics), comprehension, breadth of vocabulary, and pronunciation (including the appropriate pronunciation of loan words from English - important in Japanese). It was decided for practical reasons (the shortness of each phase), that particular criteria should be allocated as appropriate to particular phases of the test, thus politeness and comprehension were evaluated on the more interactive phases, whereas fluency and resources of grammar and expression were evaluated on those phases which were more monologic (see Appendix 2 for the assessment format).

As the assessors for the test were to be drawn not only from the language teaching field (as is the norm in language assessment) but also from the tour guiding profession (experienced guides and tour co-ordinators with guiding backgrounds), and these were found to be generally more naive and inexperienced as far as linguistic assessments are concerned[2], it was deemed necessary to develop rating scales where all the points on the scale were defined. Only by doing this was it considered possible to provide such people with a basis for discussing the appropriateness of the scores allocated during the training session, and hence to develop inter- and intra-rater reliability.

(b) Fulfilment of the task
We did not have a model for this, as we did for the linguistic criteria, of how the aspects of the assessment could be broken down. While the task specifications referred to those features in general terms which were considered to be important in overall task fulfilment (such as the ability to sympathise, to help a client make a choice, to present information clearly, to interest the listener, etc.), we now needed to refine the criteria in a way that would make them more explicit and hence ensure adequate inter- and intra-rater reliability without them being cumbersome, bearing in mind

---

[2] Even though some of the industry-based assessors are involved in selection of applicants for tour guiding positions, which involves an assessment of language ability, they tended not to have any basis for defining Japanese language performance. Assessments were generally made in a totally subjective and unverbalised manner, and it is not clear to what extent such assessments were made independently of assessments of other features.

that the assessor would also be required to conduct the interaction and that each task is approximately 4 minutes long.

The expert group was asked to rate each sample of test performance in terms of task fulfilment, and then to try to define what it was that contributed to the making of the assessment. While there was rough agreement for particular candidates on particular tasks, there was little consensus regarding the aspects of performance which were salient across all candidates, nor on the relative salience of particular features across tasks. When attempting to define *exactly* what it was that made one person more successful than another, or alternatively what was common across candidates considered to be equally successful, it appeared that any of a range of factors came into play, that lack of a particular skill in one performance might be compensated for by the presence of another, and that not all of them were equally salient for all candidates.

In trying to describe the features of successful performance, it became clear that it was impossible to make a clear distinction between language skills and other factors contributing to the assessment of performance, that whereas it was possible to separate purely linguistic features from other aspects of performance and look at them in isolation, the reverse was not possible. This is due, no doubt, to the fact that language is the medium of the performance, and is therefore integral to the success of the performance - in McNamara's words 'language is a necessary but not sufficient condition of success on the performance task'. So, for example, where a feature of the assessment was the extent to which the candidate was able to demonstrate sympathy, this was inevitably based on both the ability to express sympathy linguistically, and also on the way in which it was expressed (tone of voice, for example) and the candidate's general demeanour and responsiveness to the interlocutor throughout that task. This meant that in effect, language ability was being assessed twice - once on the linguistic criteria and again as part of the task fulfilment criteria (but in this second case the status of language vis-a-vis other non-linguistic features was undetermined). An IRT analysis of the assessment data based on the trial videos reflected this link - there was found to be a fit of linguistic assessments with task fulfilment assessments - meaning that both types of assessment were related, that language ability is a major factor in overall task fulfilment. However, looking at the raw scores we found that many students received higher or lower scores on the two types of assessment, indicating that they were not so closely related that one could be predicted from the other. It was, however, not unsurprisingly, never the case that a candidate varied widely in the two, as it is unlikely that someone with very high language ability would perform the task absolutely abysmally, or conversely that someone with very low language ability would perform the task extremely well.

Defining the task fulfilment criteria

The ease with which assessment criteria for live interactions can be applied is of importance, especially where the assessor is also the interlocutor. Thus the criteria descriptors should not be so complex that the assessor is not able to keep them in her head or refer to them quickly while conduction the interaction. Given that the balance of the features which any assessor might consider in evaluating the task fulfilment of any candidate varied, we found that the most workable approach with the task fulfilment assessments was not to define what should or should not be represented in performance (which, as mentioned, proved to be impossible anyway) let alone to define levels of skill in these features, but to to list a seris of questions which the assessor should consider in reaching this assessment (Appendix 3). These questions reflected consensus opinion on the factors the expert group generally considered relevant to the assessment of successful performance while at the same time not being so lengthy and complex that they would become unwieldy in the actual testing and assessment context.

4. The role of non-linguistic criteria in a language test.

There is an ethical question regarding the inclusion in a so-called language proficiency test assessments which reflect aspects of performance which go beyond language. Should these criteria

be included at all, and if so, to what extent? If we are interested in predicting ability to perform in specific occupational contexts through the medium of a second language (as we claim we are with all ESP tests), then we have to recognise that the predictive validity of the test is based upon the capability of the test to measure performance in similar situations and that performance is not a result of linguistic skills alone, but is affected by these other non-linguistic features. Jones (1985) addresses the issue thus:

> "With regard to second language performance testing it must be kept in mind that language is only one of several factors being evaluated. The overall criterion is the successful completion of a task in which the use of language is essential. A performance test is more than a basic proficiency test of communicative competence in that it is related to some kind of performance task. It is entirely possible for some examinees to compensate for low language proficiency by astuteness in other areas. For example, certain personality traits can assist examinees in scoring high on interpersonal tasks, even though their proficiency in the language may be substandard. On the other hand, examinees who demonstrate high general language proficiency may not score well on a performance because of deficiencies in other areas."

One might say that this has always been recognised but that assessors have had no way of addressing the issue systematically as long as the assessment criteria they have had to work with have been confined to purely linguistic features. In my own experience of being an LSP test assessor, and in discussing this role with other assessors, I have no doubt that there is a frustration with assessments which are confined to pure linguistic criteria and which do not take non-linguistic skills into account. It is often felt that, for example, particular candidates are able to compensate for poor formal skills through sheer force of personality, lack of inhibition and determination to get a point across, that they can 'communicate' but the test criteria are too narrow to allow for this. It is quite likely that many assessors of oral language have given assessments which they felt did not reflect the quality of the performance because they were limited by the criteria. Conversely, I'm sure there are cases of assessors having given higher marks than the performance warranted according to the criteria because they felt there was some *undefined* feature of the performance which made it stand out as being perticularly successful or unsuccessful communication. I am sure we all know people of whom we can say 'he/she doesn't speak the language well, but can get by very easily'.Thus the challenge is how to incorporate this overtly into the assessment model in a way that enables the assessors to give grades which reflect their overall judgement and at the same time ensuring that inter- and intra- rater reliability is maintained.

5. Conclusion

In the process of the development of the Japanese Language Test for Tour Guides a range of issues relating to the task design and the assessment of real world and linguistic criteria has to be addressed. It became clear that although there are problems associated with the definition of real-world criteria, not least because they are so specific to the particular task, where such assessment is relevant to the needs of the ultimate test users (in this case employers within the tourism industry), test developers should be responsive to these needs. The best way to ensure this relevance of assessment is to involve representatives of the industry in question at all levels of the test development process. Just because representatives of the occupational (or academic) field turn to language testers and say "Language is important in the performance we require of these people, we need help in determining who should be selected and who not", we must not allow ourselves as language testers to assume the preeminence of language in performance, to make unilateral decisions about how candidates should be tested and what the assessment criteria should be, rather we must go back to these people and find out how they themselves best consider assessments of performance should be made, what they should consist of. Their understandings and needs must be built into the test. Only by juxtaposing our insights as applied linguists with those of the

professions we purport to be serving will we succeed in producing tests which satisfy their needs as measures of performance through the second language medium.

Bibliography

Alderson, J.C. (1988) New procedures for validating proficiency tests of ESP? Theory and practice. language Testing 5,2:220-232

Bachman, L.F. (1990) *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

Bachman, L.F. (1991) What does language testing have to offer? *TESOL Quarterly* 25, 4: 671-704.

Canale; M. and M.Swain (1980) Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1,1:47.

Canale, M. (1983) On some dimensions of language proficiency in J. W. Oller (ed) *Issues in Language Testing Research.* Rowley, Mass.: Newbury House.

Douglas, D. and L. Selinker (1985) Principles for language tests within the "discourse domains" theory of interlanguage: research, test construction and interpretation. *Language Testing*, 2,2: 205-226

ITOA (1992) *ITOA Guide Accreditation Scheme: Industry discussion paper.* Inbound Tourism Organisation of Australia.

Jones, R.L. (1985) Second language performance testing: an overview in Hauptman et al (1985) *Second Language Performance Testing.* Ottawa: University of Ottawa Press.

Olshtain, E. and S. Blum-Kulka (1985) Crosscultural pragmatics and the testing of communicative competence. *Language Testing*, 2,1: 16-30

Seliger, H.W. (1985) Testing authentic language: the problem of meaning. *Language Testing*, 2,1: 1-15

Shohamy, E. and T. Reves (1985) Authentic language tests: where from and where to? *Language Testing*, 2,1: 48-59

Skehan, P. (1984) Issues in the testing of English for specific purposes. *Language Testing*, 1,2: 202-220

Spolsky, B. (1985) The limits of authenticity in language testing. *Language Testing*, 2,1: 31-40

Spolsky, B. (1985) What does it mean to know how to use a language? An essay on the theoretical basis of language testing. *Language Testing*, 2,2: 180-191

Wesche, M.B. Performance testing for work-related second language assessment, in: Shohamy, E. and R. Walton (1992) *Language Assessment for Feedback: Testing and Other Strategies,* chapter 7. Washington, D.C.:National Foreign Language Center Publications.

Appendix 1

Phase 1
The candidate is required to respond to and elaborate on questions of a personal nature and of the type that Japanese clients would be likely to ask Australian Japanese-speaking guides (such as aspects of his/her background - family, life, work, interests, home, etc.)

In items 2-6 the candidate takes on the role of a tour guide interacting with a Japanese client (the interviewer). All the items are contextualised and the purpose of the task is explained

Phase 2
The task is to help the client reach a decision about which (if any) optional tour to take on her free day. The candidate is expected to give useful advice and to encourage the client to choose a tour (out of three posibilities described in the information sheet) by answering any questions she may have and by making suggestions and promoting the tours. In this section the ability to make suggestions and offer helpful advice in a way appropriate to the situation are important.

Phase 3
The candidate is required to deal with a problem where the client is upset or worried. Examples are drawn from industry instances of 'troubleshooting' (lost property, missed plane, change in plans, illness, etc). Information is given which enables the candidate to propose a solution. The task is not to resolve the problem but to console or pacify the client while at the same time encouraging them to comply with the proposed solution.

Phase 4
The candidate is required to prepare and present a short impromptu-style talk on a culture-related topic (drawn from those in the ITOA skills module). Some background information and suggested ideas for content are given to the candidate prior to the test. They are also be encouraged to draw on their own knowledge and experience. The ability to present culture-based information relevantly, interestingly and clesrly is important.

Phase 5
Candidates are required to present detailed instructions or information to a client. Topics include the use of facilities, travel instructions, tour plans, etc. The relevant information in given in English (with key vocabulary items also given in Japanese). Candidates have some time to read through the information before presenting it. The ability to organise and present the information clearly and concisely, ensuring that the listener has understood, is important in this phase.

Phase 6
Candidates are provided with a day-tour itinerary and information on the first tourist attraction. The candidate is required to synthesise the information and prepare a short presentation. Preparation time is given prior to the test.